Request to Archive
With The National Centers for Environmental Information
For Cooling of US Midwest summer temperature extremes from cropland intensification
Provided by Harvard University

2015-10-12

This information will be used by NCEI to conduct an appraisal and make a decision on the request.

1. Who is the primary point of contact for this request?

Nathaniel Mueller
Harvard University
nmueller@fas.harvard.edu

2. Name the organization or group responsible for creating the dataset.

HARVARD/EPS > Department of Earth and Planetary Sciences, Harvard University

3. Provide an overview summarizing the scope of data you want to archive. Describe the outputs, data variables, including their measurement resolution and coverage.

This archive would document data associated with recently published manuscript in Nature Climate Change: "Cooling of US Midwest summer temperature extremes from cropland intensification" (advance online publication: October 12, 2015). Several associated items would be archived:

1. Trends in percentiles of daily maximum temperatures derived using quantile regression. These trends are for a subset of stations from the GHCN-Daily weather station data. (contained in Matlab file "RS.mat") The daily data for these stations would be included in the archive to allow reproduction of a figure in the paper.
2. Trends in summer precipitation derived for the same stations. (contained in Matlab file "RS.mat")
3. Cropland area trends derived from the gridded product of Ramankutty and Foley 1999 GBC. (associated with weather stations in the Matlab file "RS.mat", raw data by grid cell in the file "CD.mat")
4. Trends in irrigated area by county derived from digitized US agricultural census data. (associated with weather stations in the Matlab file "RS", by county in the file "IRR.mat")
5. Trends in area-normalized crop net primary productivity by county, derived from USDA annual surveys of crop areas and yields. (associated with weather stations in the file "RS.mat", by county in the file "histcropdataprocessed.mat")
6. Code relating trends in land use to trends in climate.

4. What is the time period covered by the dataset? (YYYY-MM-DD, YYYY-MM or YYYY)

From 1910 to 2014

5. Edition or version number(s) of the dataset:

1

6. Approximate date when the dataset was or will be released to the public:

2015-11

7. Who are the expected users of the archived data? How will the archived data be used?
Other scientists and students.

8. Has the dataset undergone user evaluation and/or an independent review process? Did NCEI participate in design reviews?
No

9. Describe the dataset's relationship to other archived datasets, such as earlier versions or related source data. If this is a new version, how does it improve upon the previous version(s)?
This is a new archive, although it contains data available from other sources and these are brought together for purposes of examining associations.

10. List the input datasets and ancillary information used to produce the data.
Cropland area data, an updated version of the Ramankutty and Foley 1999 GBC dataset, available at earthstat.org.

Irrigated area from recent agricultural census reports is available digitally from USDA's National Agricultural Statistics Service at http://www.nass.usda.gov/Quick_Stats/. Irrigation data prior to 1997 was digitized from agricultural census documents available at http://agcensus.mannlib.cornell.edu/AgCensus/homepage.do.

Crop area and yield data is from USDA NASS, available at http://www.nass.usda.gov/Quick_Stats/.

Conversions coefficients to derive crop NPP from yield data are largely from Monfreda et al. 2008 GBC.

11. List web pages and other links that provide information on the data.
see above

12. List the kinds of documents, metadata and code that are available for archiving. For example, data format specifications, user guides, algorithm documentation, metadata compliant with a standard such as ISO 19115, source code, platform/instrument metadata, data/process flow diagrams, etc.
1. Matlab code and data files (see answers to #15).

13. Indicate the data file format(s).
1. .m
2. .mat

14. Are the data files compressed?
gzip

15. Provide details on how the files are named and how they are organized (e.g., file_name_pattern_YYYYMM.tar in monthly aggregations).
All files and codes are contained within publicarchive.zip. This zipped folder contains:
(1) cea_paperfigures.m (code to make figures)
(2) cea_stats.m (code to analyze significance of trends)
(3) five mat files of data
- IRR.mat (irrigation data)
- CD.mat (cropland area data)
- histcropdataprocessed.mat (crop NPPan data)
- GHCNDsubsets/TMAX_1910-2014_JJA_PQC.mat (quality controlled subset of GHCND data for summer)

- fig4output/fig4data_nfe.mat (bootstrap output from a supercomputer run that contains confidence intervals for figure 4)
(4) four .mat colorbar files (designated cb_*.mat)
(5) a utils subdirectory with matlab helper functions needed to run the other codes

16. Explain how to access sample data files and/or a file listing for previewing. If it is not available now, when will it be available?
A preliminary version is available here: https://dl.dropboxusercontent.com/u/167612/publicarchive.zip

17. What is the total data volume to be submitted?
Historic Data: all historic data or data submitted as a completed collection.

Total Data Volume:	83.3MB
Number of Data Files:	1

18. Are later updates, revisions or replacement files anticipated? If so, explain the conditions for submitting these additional data to the archive.
No additional updates, revisions or replacement data are anticipated.

19. Describe the server that will connect to the ingest server at NCEI for submitting the data.

Physical Location:	This is a small file, so a simple dropbox download link should suffice.
System Name:	This is a small file, so a simple dropbox download link should suffice.
System Owner:	This is a small file, so a simple dropbox download link should suffice.
Additional Information:	

20. What are the possible methods for submitting the data to NCEI? Select all that apply.
This is a small file, so a simple dropbox download link should suffice.

21. Identify how you would like NCEI to distribute the data. Web access support depends on the resources available for the dataset.

22. Will there be any distribution, usage, or other restrictions that apply to the data in the archive?
No known constraints apply to the data.

23. Discuss the rationale for archiving the dataset and the anticipated benefits. Mention any risks associated with not archiving the dataset at NCEI.
The primary motivation is to aid in reproducible research, to allow other scholars to examine the workflow and datasets. We also anticipate that the historical irrigation dataset by county, which involved considerable hand-digitizing from pdf documents, will be of use to other researchers.

24. Are the data archived at another facility or are there plans to do so? Please explain.
No

25. Is there an existing agreement or requirement driving this request to archive? Have you already contacted someone at NCEI?
No

26. Do you have a data management plan for your data?
We have a data management plan as part of NSF Hydrologic Sciences grant 1521210. We have committed to sharing code and data online in easily usable forms.

27. Have funds been allocated to archive the data at NCEI?
No

28. Identify the affiliated research project, its sponsor, and any project/grant ID as applicable.
NSF Hydrologic Sciences grant 1521210

29. Is there a desired deadline for NCEI to archive and provide access to the data?
No deadlines for archive or access.

30. Add any other pertinent information for this request.
None